

LARGE DEVIATION RESULTS FOR MULTITYPE GALTON-WATSON TREES

BY KWABENA DOKU-AMPONSAH
University of Ghana

Abstract

Using the notion of *consistency* of empirical measures, under fairly general assumption we prove a joint large deviation principles in n for the *empirical pair measure* and *empirical offspring measure* of multitype Galton-Watson tree conditioned to have exactly n vertices in the weak topology. From these results we obtain a large deviation principle for empirical pair measure of Markov chains indexed by simply generated trees obtain by conditioning Galton-Watson trees on the total number of vertices. For the case where the offspring law of the tree is geometric distribution with parameter $\frac{1}{2}$, we get an exact rate function. All our rate functions are expressed in terms of relative entropies.

Keywords: Tree-indexed Markov chain, Tree-indexed process, random tree, Galton-Watson tree, multitype Galton-Watson tree, typed tree, joint large deviation principle, empirical pair measure, empirical offspring measure.

MSC 2000: Primary 60F10. Secondary 60J80, 05C05.

1. INTRODUCTION

Recently, conditioned Galton-Watson trees (CGWT) have received an increasing attention from researchers (see, e.g. [Al91a], [Al91b], [Al93],[DG02],[DU03],[KI11a],[KI11b],[RD11]) because of its ability to model fairly well phenomena which occur in natural hierarchy, for example, Mutations in mitochondrial DNA [OS02].

In the article [KI11a] a new type of conditioning involving the number of leaves of the Galton-Watson tree have been introduced in order to analyse a specific discrete probabilistic model, namely dissections of a regular polygon with Boltzmann weights. [RD11] dealt with the more general conditioning on having a fixed number of vertices with degree in a given set. Further, the article [KI11b] investigated the asymptotic behaviour of critical Galton-Watson trees whose offspring distribution may have infinite variance, conditioned on having a large fixed number of vertices with degree in a given set.

Large deviation analysis of Galton-Watson trees conditioned on the total size was first studied by Dembo, Mörters and Sheffield [DMS03]. In this article, notions of shift-invariance and specific relative entropy-as typically understood for Markov fields on deterministic graphs such as \mathbb{Z}^d was extended to Markov fields on random trees. With these concepts, large deviation principles for empirical measures of a class of random trees including multitype Galton-Watson trees conditioned to have exactly n vertices were proved in a topology *stronger* than the weak topology. Their analysis have shown that large deviation results, which are well-known for classical Markov chains, can be extended to Markov chains indexed by random trees with offspring laws which have superexponential decay at infinity.i.e. Offspring law $p(\cdot)$ with all its exponential moments finite.

In the course of the proof of their main result, large deviation principle for the empirical offspring measure of multitype Galton-Watson trees, see [AN72], whose exponential moments are all finite was established in their topology, see [DMS03].

However, their result failed to address offspring laws which have sub-exponential and exponential decaying function at infinity. Example, a *Markov chain indexed by geometric $\frac{1}{2}$ offspring law* which may be described as follows : First we sample a tree from a set of Galton-Watson trees with geometric distribution with parameter $\frac{1}{2}$, and then, given this tree, we run a Markov chain on the vertices of the tree in such a way that the state of a vertex depends only on the state of its parent. The result of this two-step experiment can also be interpreted as a *typed tree*.

The aim of this article is to carry out a non-trivial extension of the large deviation principle for tree indexed markov chain of [DMS03, Theorem 2.2] to cover offspring laws not discussed by [DMS03].

To be specific, we prove a joint large deviation principle for the empirical pair measure and empirical offspring measure of multitype Galton-Watson trees with offspring laws which have finite second moments. This includes offspring laws considered in [DMS03].

To deal with the problem of exponential tightness in the strong topology encountered in [DMS03] which necessitated the use of strong moment condition, we extend the concept of *consistency* as understood for empirical measures of coloured random graphs, see Doku and Mörters [DM10] or Doku [DA06], to multitype Galton-Watson trees. With this concept, we prove the upper bound, under finite second moment assumption, using the technique of (exponential) *change of measure* and two large deviations results from [DMS03, Lemma 3.6] and [DMS03, Theorem 2.2], in the weak topology.

Our proof of the lower bound unlike the upper bound uses the technique of approximating a given Multitype Galton-Watson tree from below by another Multitype Galton-Watson tree which we shall obtain by restricting the offspring distribution to some bounded set \mathcal{X}_k^* . Taking appropriate limit as k goes to infinity we obtain asymptotic results about the full tree.

Using the contraction principle, see Dembo [DZ98], we derive from our main results large deviation principle for empirical pair measure of Markov chain indexed by random trees, see Benjamini and Peres [BP94]. This result is similar to the one in [DMS03]. We remark here that the process level large deviation principles for the empirical subtree measure and single-generation empirical measure, see [DMS03], can be developed from our main results.

Specifically, we consider random tree models where trees and types are chosen *simultaneously* according to a *multitype Galton-Watson tree*. We recall from [DMS03] the model of multitype Galton-Watson tree. For a finite alphabet \mathcal{X} , we write $\mathcal{X}^* = \bigcup_{n=0}^{\infty} \{n\} \times \mathcal{X}^n$ and equip it with the discrete topology. We denote by \mathcal{T} the set of all finite rooted planar trees T , by $V = V(T)$ the set of all vertices and by $E = E(T)$ the set of all edges oriented away from the root, which is always denoted by ρ . We write $|T|$ for the number of vertices in the tree T . We note that the offspring of any vertex $v \in T$ is characterized by an element of \mathcal{X}^* and that there is an element $(0, \emptyset)$ in \mathcal{X}^* symbolizing absence of offspring. For each typed tree X and each vertex v we denote by

$$C(v) = (N(v), X_1(v), \dots, X_{N(v)}(v)) \in \mathcal{X}^*$$

the number and types of the children of v , ordered from left to right.

Given a probability measure μ on \mathcal{X} , serving as the initial distribution, and an offspring transition kernel \mathbb{Q} from \mathcal{X} to \mathcal{X}^* , we define the law \mathbb{P} of a tree-indexed process X , see Pemantle [Pe95], by the following rules:

- The root ρ carries a random type $X(\rho)$ chosen according to the probability measure μ on \mathcal{X} .

- For each vertex with type $a \in \mathcal{X}$ the offspring number and types are given independently of everything else, by the offspring law $\mathbb{Q}\{\cdot | a\}$ on \mathcal{X}^* . We write

$$\mathbb{Q}\{\cdot | a\} = \mathbb{Q}\{(N, X_1, \dots, X_N) \in \cdot | a\},$$

i.e. we have a random number N of offspring particles with types X_1, \dots, X_N .

For every $c = (n, a_1, \dots, a_n) \in \mathcal{X}^*$ and $a \in \mathcal{X}$, the *multiplicity* of the symbol a in c is given by $m(a, c) = \sum_{i=1}^n 1_{\{a_i=a\}}$, and the matrix A with index set $\mathcal{X} \times \mathcal{X}$ and nonnegative entries is given by

$$A(a, b) = \sum_{c \in \mathcal{X}^*} \mathbb{Q}\{c | b\} m(a, c), \text{ for } a, b \in \mathcal{X}. \text{ i.e. } A(a, b) \text{ are the expected number of offspring of type } a$$

of a vertex of type b . We also recall from [DMS03] the weak form of irreducibility concept. With $A^*(a, b) = \sum_{k=1}^{\infty} A^k(a, b) \in [0, \infty]$ we say that the matrix A is *weakly irreducible* if \mathcal{X} can be partitioned into a non empty set \mathcal{X}_r of *recurrent states* and a disjoint set \mathcal{X}_t of *transient states* such that

- $A^*(a, b) > 0$ whenever $b \in \mathcal{X}_r$, while
- $A^*(a, b) = 0$ whenever $b \in \mathcal{X}_t$ and either $a = b$ or $a \in \mathcal{X}_r$.

For example, any *irreducible* matrix A has A^* strictly positive, hence is also weakly irreducible with $\mathcal{X}_r = \mathcal{X}$. The multitype Galton-Watson tree is called weakly irreducible (or irreducible) if the matrix A is weakly irreducible (or irreducible, respectively) and the number $\sum_{a \in \mathcal{X}_t} m(a, c)$ of transient offspring is uniformly bounded under \mathbb{Q} .

Recall that, by the Perron-Frobenius theorem, see e.g. [DZ98, Theorem 3.1.1], the largest eigenvalue of an irreducible matrix is real and positive. Obviously, the same applies to weakly irreducible matrices. The multitype Galton-Watson tree is called *critical* if this eigenvalue is 1 for the matrix A .

The remaining part of the article is organized in the following manner: The complete statement of our results is given in Section 2, we begin with joint LDP for empirical pair measures and empirical offspring measures of multitype Galton-Watson trees, followed by a corollary of LDP for the empirical offspring measure of multitype Galton-Watson trees in subsection 2.1. In subsection 2.2 we state the LDP for empirical pair measures of Markov chains indexed by a tree. The proofs of our main results are then given in Section 3. All corollaries and Theorem 2.4 are proved in Section 4.

2. STATEMENT OF THE RESULTS

2.1 Joint large deviation principle for empirical pair measure and empirical offspring measure of multitype Galton-Watson trees.

For every sample chain X , we associate the *empirical offspring measure* M_X on $\mathcal{X} \times \mathcal{X}^*$, by

$$M_X(a, c) = \frac{1}{|T|} \sum_{v \in V} \delta_{(X(v), C(v))}(a, c), \quad (2.1)$$

and the empirical pair measure on $\mathcal{X} \times \mathcal{X}$, by

$$\tilde{L}_X(a, b) = \frac{1}{|T|} \sum_{e \in E} \delta_{(X(e_1), X(e_2))}(a, b), \text{ for } a, b \in \mathcal{X}, \quad (2.2)$$

where e_1, e_2 are the beginning and end vertex of the edge $e \in E$ (so e_1 is closer to ρ than e_2). We note that

$$\tilde{L}_X(a, b) = \sum_{c \in \mathcal{X}^*} m(b, c) M_X(a, c).$$

By definition, we notice that M_X is a probability vector and that total mass $\|\tilde{L}_X\|$ of \tilde{L}_X is $\frac{|T|-1}{|T|} \leq 1$.

Our main result is a large deviation principle for (\tilde{L}_X, M_X) if X is a multitype Galton-Watson tree.

We denote by $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ the space of probability measures ν on $\mathcal{X} \times \mathcal{X}^*$ with $\int n \nu(da, dc) < \infty$, using the convention $c = (n, a_1, \dots, a_n)$. Denote by $\tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$ the space of finite measures on $\mathcal{X} \times \mathcal{X}$ and endow the space $\tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ with the weak topology.

We call $(\varpi, \nu) \in \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ *sub-consistent* if

$$\varpi(a, b) \geq \sum_{c \in \mathcal{X}^*} m(b, c) \nu(a, c) \text{ for all } a, b \in \mathcal{X}. \quad (2.3)$$

It is call *consistent* if equality hold in (2.3). Observe that, if (ϖ, ν) is empirical pair measure and empirical offspring measure of multitype Galton-Watson tree then both sides of (2.3) is

$$\frac{1}{n} \times \sharp \{ \text{number of edges with beginning vertex of type } a \text{ and end vertex of type } b \}.$$

We call an offspring distribution \mathbb{Q} *bounded* if for some $k < \infty$, we have

$$\mathbb{Q}\{N > k \mid a\} = 0, \text{ for all } a \in \mathcal{X}.$$

Otherwise we call it *unbounded*. To formulation our first LDP, we denote by ν_1 the \mathcal{X} -marginal of probability measure ν on $\mathcal{X} \times \mathcal{X}^*$ and by ϖ_2 the second marginal of finite measure ϖ on $\mathcal{X} \times \mathcal{X}$.

Theorem 2.1. *Suppose that X is a weakly irreducible, critical multitype Galton-Watson tree with offspring law \mathbb{Q} whose second moment is finite, conditioned to have exactly n vertices. Then, for $n \rightarrow \infty$, (\tilde{L}_X, M_X) satisfies a large deviation principle in $\tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ with speed n and the convex, good rate function*

$$J(\varpi, \nu) = \begin{cases} H(\nu \parallel \nu_1 \otimes \mathbb{Q}) & \text{if } (\varpi, \nu) \text{ is sub-consistent and } \varpi_2 = \nu_1, \\ \infty & \text{otherwise,} \end{cases} \quad (2.4)$$

where ϖ_2 is the second marginal of the finite measure ϖ and ν_1 is the \mathcal{X} - marginal of the probability measure ν .

For $\nu \in \mathcal{M}(\mathcal{X} \times \mathcal{X})$, we write

$$\langle m(\cdot, c), \nu(a, c) \rangle(b) := \sum_{(a, c) \in \mathcal{X} \times \mathcal{X}^*} m(b, c) \nu(a, c), \text{ for } b \in \mathcal{X},$$

and state a corollary of Theorem 2.1.

Corollary 2.2. *Suppose that X is a weakly irreducible, critical multitype Galton-Watson tree with an offspring law \mathbb{Q} whose second moment is finite, conditioned to have exactly n vertices. Then, for $n \rightarrow \infty$, the empirical offspring measure M_X satisfies a large deviation principle in $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ with speed n and the convex, good rate function*

$$K(\nu) = \begin{cases} H(\nu \parallel \nu_1 \otimes \mathbb{Q}) & \text{if } \langle m(\cdot, c), \nu(a, c) \rangle \leq \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (2.5)$$

Here, we remark that *finite second moment* assumption in Theorem 2.1 and Corollary 2.2 is necessary for us to establish the subexponential decay of the probability of the event $\{|T| = n\}$ on the set $\{n \in \mathbb{N} : \mathbb{P}\{|T| = n\} > 0\}$. See [DMS03, Lemma 3.1].

We write $\mathcal{X}_k^* = \bigcup_{n=0}^k \{n\} \times \mathcal{X}^n$ and denote by \mathbb{Q}_k offspring transition kernel from \mathcal{X} to \mathcal{X}_k^* . The next large deviation principle is the main ingredient in the proof of the lower bound of Theorem 2.1.

Theorem 2.3. *Suppose that X is a weakly irreducible, critical multitype Galton-Watson tree with an offspring law \mathbb{Q}_k , conditioned to have exactly n vertices. Then, for $n \rightarrow \infty$, (\tilde{L}_X, M_X) satisfies a large deviation principle in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$ with speed n and the convex, good rate function*

$$J_k(\varpi, \nu) = \begin{cases} H(\nu \parallel \nu_1 \otimes \mathbb{Q}_k) & \text{if } (\varpi, \nu) \text{ is consistent and } \varpi_2 = \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (2.6)$$

2.2 LDP for empirical pair measure of Markov chains indexed by trees

In this subsection, we look at the situation where the tree is generated independently of the types.

Suppose that T is any finite tree and we are given an initial probability measure μ on a finite alphabet \mathcal{X} and a Markovian transition kernel $Q: \mathcal{X} \times \mathcal{X} \geq 0$. We can obtain a *Markov chain indexed by tree* T , $X: V \rightarrow \mathcal{X}$ as follows: Choose $X(\rho)$ according to μ and choose $X(v)$, for each vertex $v \neq \rho$, using the transition kernel given the value of its parent, independently of everything else. If the tree is chosen randomly, we always consider $X = \{X(v) : v \in T\}$ under the *joint law* of tree and chain. It is sometimes convenient to interpret X as a *typed tree*, considering $X(v)$ as the *type* of the vertex v .

We consider the class of *simply generated trees*, see [MM78] or [Al91a], obtained by conditioning a critical Galton-Watson on its total number of vertices. To be specific, we look at the class of Galton-Watson trees, where the number of children $N(v)$ of each $v \in T$ is chosen independently according to the same law $p(\cdot) = \mathbb{P}\{N(v) = \cdot\}$ for all $v \in T$, while $0 < p(0) < 1$. We assume that p is critical. That is, the mean offspring number $\sum_{\ell=0}^{\infty} \ell p(\ell)$ is one, but this assumption is not restrictive: Note that the distribution of T conditioned on $\{|T| = n\}$ is exactly the same as when the offspring law is $p_\theta(\ell) = p(\ell)e^{\theta\ell} / \sum_j p(j)e^{\theta j}$, regardless of the value of $\theta \in \mathbb{R}$. With $0 < p(0) < 1 - p(1)$ there exists a unique θ_* such that $\sum_{\ell} \ell p_{\theta_*}(\ell) = 1$. Hence all our results hold in the noncritical cases with p_{θ_*} in place of p . We allow offspring laws p with unbounded support, but we relax the assumption the all exponential moments of p are finite. i.e. we relax the assumption $\ell^{-1} \log p(\ell) \rightarrow -\infty$.

We shall assume hereafter that the statement conditioned on the event $\{|T| = n\}$ are made only for those values of n where the event $\{|T| = n\}$ has positive probability.

For each typed tree X , we recall from [DMS03], the definition of *empirical pair (probability) measure* L_X on $\mathcal{X} \times \mathcal{X}$ as

$$L_X(a, b) = \frac{1}{|E|} \sum_{e \in E} \delta_{(X(e_1), X(e_2))}(a, b), \quad \text{for } a, b \in \mathcal{X}, \quad (2.7)$$

where e_1, e_2 are the beginning and end vertex of the edge $e \in E$ (so e_1 is closer to ρ than e_2). Notice, $L_X = \frac{n}{n-1} \tilde{L}_X$ on the set $\{|T| = n\}$ and hence the LDP for \tilde{L}_X implies L_X by exponential equivalent Theorem, see Dembo [DZ98]. Our first result in this subsection is a large deviation principle for L_X .

Theorem 2.4. *Suppose that T is a Galton-Watson tree, with offspring law $p(\cdot)$ such that $0 < p(0) < 1 - p(1)$, $\sum_{\ell} \ell p(\ell) = 1$ and $\sum_{\ell} \ell^2 p(\ell) < \infty$. Let X be a Markov chain indexed by T with arbitrary initial distribution and an irreducible Markovian transition kernel Q . Then, for $n \rightarrow \infty$, the empirical pair measure L_X , conditioned on $\{|T| = n\}$ satisfies a large deviation principle in $\mathcal{M}(\mathcal{X} \times \mathcal{X})$ with speed n and the convex, good rate function*

$$I(\mu) = \begin{cases} H(\mu \parallel \mu_1 \otimes Q) + \sum_{a \in \mathcal{X}} \mu_2(a) I_p\left(\frac{\mu_1(a)}{\mu_2(a)}\right) & \text{if } \mu_1 \ll \mu_2, \\ \infty & \text{otherwise,} \end{cases} \quad (2.8)$$

where

$$I_p(x) = \sup_{\lambda \in \mathbb{R}} \left\{ \lambda x - \log \left[\sum_{n=0}^{\infty} p(n) e^{\lambda n} \right] \right\}, \quad (2.9)$$

and μ_1, μ_2 are the first and second marginal of μ respectively, and $\mu_1 \otimes Q(a, b) = Q\{b | a\} \mu_1(a)$.

From Theorem 2.4 we obtain LDP for empirical pair measures of Galton-Watson trees with geometric distribution with parameter $\frac{1}{2}$.

Corollary 2.5. *Suppose that T is a Galton-Watson tree, with offspring law $p(\ell) = 2^{-(\ell+1)}$, $\ell = 0, 1, \dots$. Let X be a Markov chain indexed by T with arbitrary initial distribution and an irreducible Markovian transition kernel Q . Then, for $n \rightarrow \infty$, the empirical pair measure L_X , conditioned on $\{|T| = n\}$ satisfies a large deviation principle in $\mathcal{M}(\mathcal{X} \times \mathcal{X})$ with speed n and the convex, good rate function*

$$I(\mu) = \begin{cases} H(\mu \| \mu_1 \otimes Q) + H(\mu_1 \| (\mu_1 + \mu_2)/2) + H(\mu_2 \| (\mu_1 + \mu_2)/2) & \text{if } \mu_1 \ll \mu_2, \\ \infty & \text{otherwise,} \end{cases} \quad (2.10)$$

where μ_1 and μ_2 are the first and second marginal of μ and $\mu_1 \otimes Q(a, b) = Q\{b | a\} \mu_1(a)$.

3. PROOF OF MAIN RESULTS

3.1 Change of Measure, Exponential Tightness and Some General Principles.

Denote by \mathcal{C} the space of bounded functions on $\mathcal{X} \times \mathcal{X}^*$ and for $\tilde{g} \in \mathcal{C}$, we define the function

$$U_{\tilde{g}}(a) = \log \sum_{c \in \mathcal{X}^*} e^{\tilde{g}(a, c)} \mathbb{Q}\{c | a\}, \quad (3.1)$$

for $a \in \mathcal{X}$. Using \tilde{g} we define the following new multitype Galton-Watson tree :

- Assign the root ρ , type $a \in \mathcal{X}$ according to the probability distribution $\mu_{\tilde{g}}(a)$ given by

$$\mu_{\tilde{g}}(a) = \frac{\mu(a) e^{U_{\tilde{g}}(a)}}{\int \mu(db) e^{U_{\tilde{g}}(b)}}. \quad (3.2)$$

- For every vertex with type $a \in \mathcal{X}$ the offspring number and types are given independently of everything else, by the offspring law $\tilde{\mathbb{Q}}\{\cdot | a\}$ given by

$$\tilde{\mathbb{Q}}\{c | a\} = \mathbb{Q}\{c | a\} \exp(\tilde{g}(a, c) - U_{\tilde{g}}(a)). \quad (3.3)$$

By $\tilde{\mathbb{P}}$ we denote the transformed law and observe that $\tilde{\mathbb{P}}$ is absolutely continuous with respect to \mathbb{P} . Specifically, for each finite X ,

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}(X) = \frac{e^{U_{\tilde{g}}(X(\rho))}}{\int e^{U_{\tilde{g}}(b)} \mu(db)} \prod_{v \in V} \exp \left[\tilde{g}(X(v), C(v)) - U_{\tilde{g}}(X(v)) \right] \quad (3.4)$$

$$= \frac{1}{\int e^{U_{\tilde{g}}(a)} \mu(da)} \prod_{v \in V} \exp \left[\tilde{g}(X(v), C(v)) - \sum_{b \in \mathcal{X}} m(b, C(v)) U_{\tilde{g}}(b) \right] \quad (3.5)$$

$$= \frac{1}{\int e^{U_{\tilde{g}}(a)} \mu(da)} \exp \left[\langle \tilde{g} - \sum_{b \in \mathcal{X}} m(b, \cdot) U_{\tilde{g}}(b), M_X \rangle \right], \quad (3.6)$$

where $C(v) = (N(v), X_1(v), \dots, X_N(v))$. We recall from [DMS03] the following exponential tightness theorem of the family of laws of M_X on the space $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ equipped with their topology.

Lemma 3.1. *For every $\alpha > 0$ there exists a compact $K_\alpha \subset \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ with*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{M_X \notin K_\alpha \mid |T| = n\} \leq -\alpha.$$

Let $l \in \mathbb{N}$, and $B(l) \in \mathbb{N}$ large enough such that $\mathbb{Q}\{e^{l^2 \mathbf{1}_{\{N > B(l)\}}}\} \leq 2^l$, for all a . Using the exponential chebyshev's inequality we obtain,

$$\begin{aligned} \mathbb{P}\{M_X[N > B(l)] \geq l^{-1}, |T| = n\} &\leq e^{-nl} \mathbb{E}\left\{e^{l^2 \sum_{v \in V} \mathbf{1}_{\{N(v) > B(l)\}}}, |T| = n\right\} \\ &= e^{-ln} \mathbb{E}\left\{\prod_{v \in T} \exp(l^2 \mathbf{1}_{\{N(v) > B(l)\}}), |T| = n\right\} \\ &\leq e^{-nl} \left(\sup_{a \in \mathcal{X}} \mathbb{Q}\{\exp(l^2 \mathbf{1}_{\{N > B(l)\}}) \mid a\}\right)^n \leq e^{-n(l - \log 2)}. \end{aligned}$$

Fix α and choose $M > \alpha + \log 2$. Define the set $\Gamma_M = \{\nu : \nu[N > B(l)] < l^{-1}, l \geq M\}$. Note, $\{N \leq B(l)\} \subset \mathcal{X} \times \mathcal{X}^*$ is compact, and so we have that the set Γ_M is pre-compact in the weak topology, by Prohorov's criterion. Moreover,

$$\mathbb{P}\{M_X \notin \Gamma_M \mid |V(T)| = n\} \leq \frac{1}{\mathbb{P}\{|T| = n\}} \frac{1}{1 - e^{-1}} \exp(-n(M - \log 2)),$$

hence using [DMS03, Lemma 3.1] we get

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{M_X \notin K_\alpha \mid |V(T)| = n\} \leq -\alpha,$$

for the closure K_α of Γ_M . This ends the proof of the tightness Lemma.

We remark here that this result also holds for $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ endowed with the weak topology.

We denote by \mathcal{M}_s the set of all sub-consistent measures, and by \mathcal{M}_c the set of all consistent measures in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ and notice that $\mathcal{M}_c \subseteq \mathcal{M}_s$. For k a natural number, we denote by $\mathcal{M}_{c,k}$ the set of consistent measures in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$. Then, \mathcal{M}_s is a closed subset of $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ and $\mathcal{M}_{c,k}$ is a closed subset of $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$. The next two large deviation principles will help us extend LDP in $\mathcal{M}_{c,k}$, \mathcal{M}_s to $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$ and $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ respectively.

Lemma 3.2. *Suppose X is a multitype Galton-Watson tree with offspring law \mathbb{Q} . Assume (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ satisfies the LDP in \mathcal{M}_s with convex, good rate function*

$$\tilde{J}(\varpi, \nu) = \begin{cases} H(\nu \parallel \nu_1 \otimes \mathbb{Q}) & \text{if } \varpi_2 = \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (3.7)$$

Then, (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ satisfies the LDP in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^)$ with convex, good rate function J .*

Proof. Observe that, $\{|T| = n\} := \{\omega : |T|(\omega) = n\} \subseteq \{\omega : (\tilde{L}_X, M_X)(\omega) \in \mathcal{M}_c\} =: \{(\tilde{L}_X, M_X) \in \mathcal{M}_c\}$ and so, for all n , we have $\mathbb{P}\{(\tilde{L}_X, M_X) \in \mathcal{M}_s \mid |T| = n\} = 1$. Also, if $(\varpi_n, \nu_n) \in \mathcal{M}_s$ converges to (ϖ, ν) then by the Fatou's Lemma, we have that

$$\varpi(a, b) = \lim_{n \rightarrow \infty} \varpi_n(a, b) \geq \lim_{n \rightarrow \infty} \sum_{c \in \mathcal{X}^*} m(b, c) \nu_n(a, c) \geq \liminf_{n \rightarrow \infty} \sum_{c \in \mathcal{X}^*} m(b, c) \nu_n(a, c) \geq \sum_{c \in \mathcal{X}^*} m(b, c) \nu(a, c),$$

which implies (ϖ, ν) is sub-consistent. This means \mathcal{M}_s is a closed subset of $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$. Therefore, by [DZ98, Lemma 4.1.5], the LDP for (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ holds with convex, good rate function J . \square

Lemma 3.3. *Suppose X is a multitype Galton-Watson tree with offspring law \mathbb{Q}_k . Assume (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ satisfies the LDP in $\mathcal{M}_{c,k}$ with convex, good rate function*

$$\tilde{J}_k(\varpi, \nu) = \begin{cases} H(\nu \parallel \nu_1 \otimes \mathbb{Q}_k) & \text{if } \varpi_2 = \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (3.8)$$

Then, (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ satisfies the LDP in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^)$ with convex, good rate function J_k .*

Proof. Using the same argument as in the proof of Lemma 3.2 we have $\mathbb{P}\{(\tilde{L}_X, M_X) \in \mathcal{M}_{c,k} \mid |T| = n\} = 1$. Moreover, as $m(a, c) \leq k$ for all $(a, c) \in \mathcal{X} \times \mathcal{X}_k^*$ if $(\varpi_n, \nu_n) \in \mathcal{M}_{c,k}$ converges (ϖ, ν) then we have

$$\varpi(a, b) = \lim_{n \rightarrow \infty} \varpi_n(a, b) = \lim_{n \rightarrow \infty} \sum_{c \in \mathcal{X}_k^*} m(b, c) \nu_n(a, c) = \sum_{c \in \mathcal{X}_k^*} m(b, c) \nu(a, c),$$

which implies (ϖ, ν) is consistent. This means $\mathcal{M}_{c,k}$ is a closed subset of $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$. Hence, by [DZ98, Lemma 4.1.5], the LDP for (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ holds with convex, good rate function J_k which completes the proof of the Lemma. \square

In view of Lemmas 3.3 and 3.2, we establish large deviation principles in the spaces $\mathcal{M}_{c,k}$ and \mathcal{M}_s .

3.2 Proof of the upper bound in Theorem 2.1. Next we derive an upper bound in a variational formulation. Denote by \mathcal{C} the space of bounded functions on $\mathcal{X} \times \mathcal{X}^*$ and define for each (ϖ, ν) sub-consistent element in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$, the function \hat{J} by

$$\begin{aligned} \hat{J}(\varpi, \nu) &= \sup_{g \in \mathcal{C}} \left\{ \int g(b, c) \nu(db, dc) - \int U_g(b) \varpi(da, db) \right\} \\ &= \sup_{g \in \mathcal{C}} \left\{ \int g(b, c) \nu(db, dc) - \int \sum_{b \in \mathcal{X}} m(b, c) U_g(b) \nu(da, dc) \right\} \end{aligned}$$

where $c = (n, a_1, \dots, a_n)$.

Lemma 3.4. *For each closed set $F \subset \mathcal{M}_s$, we have*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{(\tilde{L}_X, M_X) \in F \mid |T| = n\} \leq - \inf_{(\varpi, \nu) \in F} \hat{J}(\varpi, \nu).$$

Let $\tilde{g} \in \mathcal{C}$ be bounded by M . Note from the definition of $U_{\tilde{g}}$ from (3.1) that $U_{\tilde{g}} \leq M$. Using (3.4), we obtain

$$e^M \geq \int e^{U_{\tilde{g}}(a)} \mu(da) = \mathbb{E} \left\{ \exp \left[\langle \tilde{g} - \sum_{b \in \mathcal{X}} m(b, \cdot) U_{\tilde{g}}(b), M_X \rangle \right], |V(T)| = n \right\}$$

Now, we take limit as n approaches infinity and use [DMS03, Lemma 3.1] to obtain

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left\{ \exp \left[\langle \tilde{g} - \sum_{b \in \mathcal{X}} m(b, \cdot) U_{\tilde{g}}(b), M_X \rangle \right] \mid |V(T)| = n \right\} \leq 0 \quad (3.9)$$

Similarly, we can use (3.4) and [DMS03, Lemma 3.1] to obtain

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left\{ \exp [\langle \tilde{g} - U_{\tilde{g}}, M_X \rangle] \mid |V(T)| = n \right\} \leq 0. \quad (3.10)$$

Next, we write $J_\varepsilon(\varpi, \nu) := \min\{J(\varpi, \nu), \varepsilon^{-1}\} - \varepsilon$. Fix $(\varpi, \nu) \in F$ and choice $\tilde{g} \in \mathcal{C}$, such that

$$[\langle \tilde{g}, \nu \rangle - \langle U_{\tilde{g}}, \varpi \rangle] \geq J_\varepsilon(\varpi, \nu)$$

Now, since \tilde{g} and $U_{\tilde{g}}$ are both bounded function we can find open neighbourhood B_ϖ , B_ν of ϖ and ν respectively, such that we have

$$\inf_{\substack{\tilde{\varpi} \in B_\varpi \\ \tilde{\nu} \in B_\nu}} [\langle \tilde{g}, \tilde{\nu} \rangle - \langle U_{\tilde{g}}, \tilde{\varpi} \rangle] \geq [\langle \tilde{g}, \nu \rangle - \langle U_{\tilde{g}}, \varpi \rangle] \geq J_\varepsilon(\varpi, \nu) - \varepsilon \quad (3.11)$$

Applying the exponential Chebyshev inequality to (3.11) and using (3.9) we obtain that,

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \{ (\tilde{L}_X, M_X) \in B_\varpi \times B_\nu \mid |V(T)| = n \} \\ & \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left\{ \exp [\langle \tilde{g} - \sum_{b \in \mathcal{X}} m(b, \cdot) U_{\tilde{g}}(b), M_X \rangle] \mid |V(T)| = n \right\} - J_\varepsilon(\varpi, \nu) + \varepsilon \\ & \leq - \inf_{(\varpi, \nu) \in F} J_\varepsilon(\varpi, \nu) + \varepsilon. \end{aligned} \quad (3.12)$$

Let assume that (ϖ, ν) fails to be sub-consistent i.e. There exists $a \in \mathcal{X}$ such that

$$\nu_1(a) < \varpi_2(a). \quad (3.13)$$

Then we can find $\delta > 0$ and a small open neighbourhood $B_\varpi \times B_\nu \subset \mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ such that

$$\tilde{\nu}_1(a) < \tilde{\varpi}_2(a) - \delta, \text{ for all } (\tilde{\varpi}, \tilde{\nu}) \in B_\varpi \times B_\nu. \quad (3.14)$$

We write $\tilde{g}(b, c) = -(\delta\varepsilon)^{-1} \mathbf{1}_a(b)$ and note that, by the definition (3.1), we have $U_{\tilde{g}}(b) = \tilde{g}(b, c)$ for all b and this vanishes unless $b = a$. Hence, by (3.14), for every $(\tilde{\varpi}, \tilde{\nu}) \in B_\varpi \times B_\nu$ we have that

$$\int (\tilde{g} - \sum_{b \in \mathcal{X}} m(b, \cdot) U_{\tilde{g}}(b)) d\tilde{\nu} > \varepsilon^{-1}.$$

We use the exponential Chebyshev inequality and (3.11) to obtain

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \{ (\tilde{L}_X, M_X) \in B_\varpi \times B_\nu \mid |V(T)| = n \} \\ & \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left\{ \exp [\langle \tilde{g} - \sum_{b \in \mathcal{X}} m(b, \cdot) U_{\tilde{g}}(b), M_X \rangle] \mid |V(T)| = n \right\} - \varepsilon^{-1} \leq -\varepsilon^{-1} \\ & \leq - \inf_{(\varpi, \nu) \in F} J_\varepsilon(\varpi, \nu) + \varepsilon. \end{aligned} \quad (3.15)$$

Now we use Lemma 3.1 to choose a compact set K_α (for $\alpha = \varepsilon^{-1}$) with

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \{ M_X \notin K_\alpha \mid |V(T)| = n \} \leq -\varepsilon^{-1}. \quad (3.16)$$

For this K_α we denote by

$$\Gamma_\alpha := \{(\varpi, \nu) : (\varpi, \nu) \in \mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*), \nu \in K_\alpha\}.$$

The set $\Gamma_\alpha \cap F$ is compact and hence it may be covered by finitely many of the sets $B_{\varpi_1} \times B_{\nu_1}, \dots, B_{\varpi_m} \times B_{\nu_m}$, with $(\varpi_i, \nu_i) \in F$ for $i = 1, \dots, m$. Hence,

$$\mathbb{P}\{M_X \in F \mid |T| = n\} \leq \sum_{i=1}^m \mathbb{P}\{M_X \in B_{\nu_i} \mid |T| = n\} + \mathbb{P}\{M_X \notin K_\alpha \mid |T| = n\}.$$

Using (3.12) and (3.15) we obtain, for small enough $\varepsilon > 0$, that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{(\tilde{L}_X, M_X) \in F \mid |T| = n\} &\leq \max_{i=1}^m \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{(\tilde{L}_X, M_X) \in B_{\varpi_i} \times B_{\nu_i} \mid |V(T)| = n\} \\ &\leq - \inf_{(\varpi, \nu) \in F} \hat{J}_\varepsilon(\varpi, \nu) + \varepsilon. \end{aligned}$$

Taking $\varepsilon \downarrow 0$ gives the required statement

lemma is completely analogous to the proof of [DMS03, Lemma 3.3 ,p. 983].

Recall that $\tilde{J} : \mathcal{M}_s \rightarrow [0, \infty]$ is given by

$$\tilde{J}(\varpi, \nu) = \begin{cases} H(\nu \parallel \nu_1 \otimes \mathbb{Q}) & \text{if } \varpi_2 = \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (3.17)$$

We show that the convex rate function \tilde{J} may replace the function \hat{J} of (3.2) in the upper bound of Lemma 3.4.

Lemma 3.5. *The function \tilde{J} is convex and lower semicontinuous on \mathcal{M}_s . Moreover, $\tilde{J}(\varpi, \nu) \leq \hat{J}(\varpi, \nu)$, for any $(\varpi, \nu) \in \mathcal{M}_s$.*

The proof of the inequality $\tilde{J}(\varpi, \nu) \leq \hat{J}(\varpi, \nu)$ is analogous to the proof of [DMS03, Lemma 3.4]. To prove that \tilde{J} is convex, good rate function, we consider the convex, good rate function $\phi : \mathbb{R} \rightarrow [0, \infty]$ given by $\phi(x) = x \log x - x + 1$. Then, we can represent the left side of (3.17) in the form

$$H(\nu \parallel \nu_1 \otimes \mathbb{Q}) = \begin{cases} \int \phi \circ f d(\nu_1 \otimes \mathbb{Q}) & \text{if } f := \frac{d\nu}{d(\nu_1 \otimes \mathbb{Q})} \text{ exists} \\ \infty & \text{otherwise.} \end{cases} \quad (3.18)$$

Consequently, by [DZ98, Lemma 6.2.16], \tilde{J} is a convex, good rate function.

By Lemma 3.2 the large deviation upper bound Lemma 3.4 holds with rate function \tilde{J} replaced by J .

3.3 Proof of Theorem 2.3. We begin by recalling from [DMS03] that $\nu \in \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ is shift-invariant if

$$\nu_1(a) = \sum_{(b,c) \in \mathcal{X} \times \mathcal{X}^*} m(a, c) \nu(b, c), \text{ for all } a \in \mathcal{X}.$$

This Theorem is derived from [DMS03, Theorem 2.2] by applying the contraction principle to the linear mapping $G : \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*) \mapsto \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$ given by $G(\nu) = (\varpi, \nu)$ where (ϖ, ν) is consistent. Specifically, [DMS03, Theorem 2.2] implies the large deviation for $G(M_X) = (\tilde{L}, M_X)$ with convex, good rate function $\tilde{J}_k(\varpi, \nu) = \inf \{K_k(\nu) : G(\nu) = (\varpi, \nu), (\varpi, \nu) \text{ is consistent}\}$, where

$$K_k(\nu) = \begin{cases} H(\nu \parallel \nu_1 \otimes \mathbb{Q}_k) & \text{if } \langle m(\cdot, c), \nu(a, c) \rangle = \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (3.19)$$

Using shift-invariance and consistency we have

$$\nu_1(a) = \sum_{(b,c) \in \mathcal{X} \times \mathcal{X}_k^*} m(a, c) \nu(b, c) = \sum_{b \in \mathcal{X}} \varpi(b, a) = \varpi_2(a), \text{ for all } a \in \mathcal{X}.$$

Therefore, by Lemma 3.3, the LDP for (\tilde{L}, M_X) conditional on the event $\{|T| = n\}$ holds in $\tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$ with convex, good rate function J_k .

3.4 Proof of the Lower Bound in Theorem 2.1

We define for every weakly irreducible, critical offspring kernel $\mathbb{Q}\{c|b\}$ the conditional offspring law

$$\mathbb{Q}_k\{c|a\} = \begin{cases} \frac{1}{\mathbb{Q}\{\mathcal{X}_k^*|a\}} \mathbb{Q}\{c|a\} & \text{if } c \in \mathcal{X}_k^* \\ 0 & \text{otherwise,} \end{cases} \quad (3.20)$$

where $\mathbb{Q}\{\mathcal{X}_k^*|a\} = \sum_{c \in \mathcal{X}_k^*} \mathbb{Q}\{c|a\}$. We write $\|\nu\|_k = \nu(\mathcal{X} \times \mathcal{X}_k^*)$ and observe that

$$\lim_{k \rightarrow \infty} \nu(\mathcal{X} \times \mathcal{X}_k^*) = \lim_{k \rightarrow \infty} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*} \mathbf{1}_{\{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*\}} \nu(a,c) = 1,$$

by the dominated convergence. We define ν_k a probability measure on $\mathcal{X} \times \mathcal{X}^*$ by

$$\nu_k(a,c) = \begin{cases} \frac{\nu(a,c)}{\|\nu\|_k} & \text{if } (a,c) \in \mathcal{X} \times \mathcal{X}_k^* \\ 0 & \text{otherwise,} \end{cases} \quad (3.21)$$

and denote by $(\nu_k)_1$ the \mathcal{X} -marginal of the probability measure ν_k . We write

$$\varpi_k(a,b) := \sum_{c \in \mathcal{X}_k^*} m(b,c) \nu_k(a,c).$$

Note that for any bounded function $f : \mathcal{X} \times \mathcal{X}^* \rightarrow \mathbb{R}$, we have $\lim_{k \rightarrow \infty} \int f d\nu_k \rightarrow \int f d\nu$. We write for $(a,c) \in \mathcal{X} \times \mathcal{X}^*$

$$f_k(a,c) := \mathbf{1}_{\{\nu_k(a,c) > 0\}} \log \frac{\nu_k(a,c)}{(\nu_k)_1 \otimes \mathbb{Q}_k(a,c)},$$

$$f^{(0,q]}(a,c) := \mathbf{1}_{\{\nu_1 \otimes \mathbb{Q}(a,c) \geq \nu(a,c) > 0\}} \log \frac{\nu(a,c)}{\nu_1 \otimes \mathbb{Q}(a,c)},$$

$$f^{(q,1]}(a,c) := \mathbf{1}_{\{1 \geq \nu(a,c) > \nu_1 \otimes \mathbb{Q}(a,c)\}} \log \frac{\nu_k(a,c)}{\nu_1 \otimes \mathbb{Q}(a,c)}$$

and observe that $f^{(q,1]}$ positive bounded function but $f^{(0,q]}$ is non-positive, unbounded continuous function.

Lemma 3.6 (Limit Inequalities). *Let $\nu \ll \nu_1 \otimes \mathbb{Q}$. Then, the following limits holds:*

- (i) $\limsup_{k \rightarrow \infty} \langle f^{(0,q]}, \nu_k \rangle \leq \langle f^{(0,q]}, \nu \rangle,$
- (ii) $\limsup_{k \rightarrow \infty} \langle f_k, \nu_k \rangle \leq \langle f^{(0,q]}, \nu \rangle + \langle f^{(q,1]}, \nu \rangle.$

Proof. (i) Recall that we have assumed $\nu \ll \nu_1 \otimes \mathbb{Q}$ and note that by definition $f^{(0,q]} \leq 0$. Using [DZ98, Theorem D.12, p.357], we obtain

$$\liminf_{k \rightarrow \infty} \langle -f^{(0,q]}, \nu_k \rangle \geq \langle -f^{(0,q]}, \nu \rangle.$$

Factorizing -1 out of the integrals and divide both side of the inequality by -1 , we obtain the limit inequality (i) of Lemma (3.6).

(ii) Notice, $\lim_{k \rightarrow \infty} f_k = f^{(0,q]} + f^{(q,1]}$. Hence, for any $\delta > 0$ and sufficiently large k we have that

$$\langle f_k, \nu_k \rangle \leq \langle f^{(0,q]}, \nu_k \rangle + \langle f^{(q,1]}, \nu_k \rangle + \delta. \quad (3.22)$$

Next, we take limit as k approaches ∞ of all sides of (3.22) and Lemma (3.6) (i) to obtain

$$\limsup_{k \rightarrow \infty} \langle f_k, \nu_k \rangle \leq \langle f^{(0,q]}, \nu \rangle + \langle f^{(q,1]}, \nu \rangle + \delta. \quad (3.23)$$

Taking $\delta \downarrow 0$ we have the desire result which concludes the proof of the Lemma. \square

We define the total variation metric d by

$$d(\nu, \tilde{\nu}) = \frac{1}{2} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} |\nu(a, c) - \tilde{\nu}(a, c)|. \quad (3.24)$$

This metric generates the weak topology. By $\mathcal{N}(\mathcal{X})$ we denote the space of counting measures on \mathcal{X} . We recall that \mathcal{M}_s denotes set of sub-consistent measures in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$, $\mathcal{M}_{c,k}$ denotes set of consistent measures in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$. In the next five Lemmas, we approximate $J(\varpi, \nu)$ for a sub-consistent pair (ϖ, ν) by $J_k(\varpi_k, \nu_k)$ with (ϖ_k, ν_k) consistent.

Lemma 3.7 (Approximation Step 1). *Suppose (ϖ, ν) is sub-consistent. Then, there exists a sequence $(\hat{\varpi}_n, \hat{\nu}_n) \in \mathcal{M}_s$ such that*

- (i) $(\hat{\varpi}_n, \hat{\nu}_n)$ is consistent.
- (ii) $(\hat{\varpi}_n, \hat{\nu}_n) \rightarrow (\varpi, \nu)$ as $n \rightarrow \infty$.

Proof. For any $b \in \mathcal{X}$ we define $e^{(b)} \in \mathcal{N}(\mathcal{X})$ by $e^{(b)}(a) = 0$ if $a \neq b$, and $e^{(b)}(a) = 1$ if $a = b$. We write $m(c) = (m(a, c), a \in \mathcal{X})$ and for large n define $\hat{\nu}_n \in \mathcal{M}_s$ by

$$\hat{\nu}_n(a, c) = \nu(a, c) \left(1 - \frac{\|\varpi - \langle m(\cdot, c), \nu(\cdot, c) \rangle\|}{n} \right) + \sum_{b \in \mathcal{X}} 1\{m(c) = ne^{(b)}\} \frac{\varpi(a, b) - \langle m(\cdot, c), \nu(\cdot, c) \rangle(a, b)}{n}. \quad (3.25)$$

We note that $\hat{\nu}_n \rightarrow \nu$ and that, for all $a, b \in \mathcal{X}$,

$$\begin{aligned} \sum_{c \in \mathcal{X}^*} m(a, c) \hat{\nu}_n(b, c) &= \left(1 - \frac{\|\varpi - \langle m(\cdot, c), \nu(\cdot, c) \rangle\|}{n} \right) \sum_{c \in \mathcal{X}^*} m(a, c) \nu(b, c) + \varpi(a, b) - \langle m(\cdot, c), \nu(\cdot, c) \rangle(a, b) \\ &= \varpi(a, b) - \frac{\|\varpi - \langle m(\cdot, c), \nu(\cdot, c) \rangle\|}{n} \langle m(\cdot, c), \nu(\cdot, c) \rangle(a, b) \xrightarrow{n \uparrow \infty} \varpi(a, b). \end{aligned}$$

Defining $\hat{\varpi}_n$ by $\hat{\varpi}_n(a, b) = \sum_{c \in \mathcal{X}^*} m(a, c) \hat{\nu}_n(b, c)$, we have a sequence of consistent pairs $(\hat{\varpi}_n, \hat{\nu}_n)$ converging to (ϖ, ν) . \square

From $(\hat{\varpi}_n, \hat{\nu}_n)$ we construct a sequence of $(\hat{\varpi}_{k,n}, \hat{\nu}_{k,n})$ in $\mathcal{M}_{c,k}$ closed in limit to (ϖ, ν) . To this end, for $k \in \mathbb{N}$, we define $\hat{\nu}_{k,n}$ a probability measure on $\mathcal{X} \times \mathcal{X}^*$ by

$$\hat{\nu}_{k,n}(a, c) = \begin{cases} \frac{\hat{\nu}_n(a, c)}{\|\hat{\nu}_n\|_k} & \text{if } (a, c) \in \mathcal{X} \times \mathcal{X}_k^* \\ 0 & \text{otherwise,} \end{cases} \quad (3.26)$$

and denote by $(\hat{\nu}_{k,n})_1$ the \mathcal{X} -marginal of the probability measure $\hat{\nu}_{k,n}$. We take

$$\hat{\varpi}_{k,n}(a, b) := \sum_{c \in \mathcal{X}_k^*} m(b, c) \hat{\nu}_{k,n}(a, c).$$

Lemma 3.8 (Approximation Step 2). *Suppose (ϖ, ν) is sub-consistent. Then, for every $\varepsilon > 0$, there exists $k(\varepsilon)$ and $n(\varepsilon)$ such that for large $k > k(\varepsilon)$, $n > n(\varepsilon)$, we have*

$$|\varpi(a, b) - \varpi_{k,n}(a, b)| < \varepsilon, \text{ for all } a, b \in \mathcal{X},$$

Proof. Let $\varepsilon > 0$ and take $\delta = \frac{\varepsilon}{3}$. We choose $k(\varepsilon)$ large enough such that

$$\sup_{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*} \left| \frac{1_{\{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*\}}}{\|\hat{\nu}_n\|_k} - 1 \right| \leq \delta.$$

For $k \geq k(\varepsilon)$, we recall the definition $\hat{\nu}_{k,n}$ and $\hat{\varpi}_{k,n}$, and observe that

$$\begin{aligned} d(\hat{\nu}_{k,n}, \hat{\nu}_n) &= \frac{1}{2} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} |\hat{\nu}_{k,n}(a, c) - \hat{\nu}_n(a, c)| = \frac{1}{2} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \hat{\nu}_n(a, c) \left| \frac{1_{\{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*\}}}{\|\hat{\nu}_n\|_k} - 1 \right| \leq \frac{1}{2} \delta \\ &= \varepsilon/3 \end{aligned}$$

Define for large $n > 1$ the measure $\tilde{\varpi}_n$ by $\tilde{\varpi}_n(a, b) = (1 + \frac{1}{n})\varpi(a, b)$, for $a, b \in \mathcal{X}$. Then, it is not hard to see that

$$\lim_{n \rightarrow \infty} \tilde{\varpi}_n(a, b) = \varpi(a, b), \text{ for all } a, b \in \mathcal{X}.$$

i.e. for every $\varepsilon > 0$ there exists $n(\varepsilon) \in \mathbb{N}$ such that for all $n > n(\varepsilon)$ implies $|\tilde{\varpi}_n(a, b) - \varpi(a, b)| \leq \varepsilon/2$. Now, observe we have

$$\lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \hat{\varpi}_{k,n}(a, b) \geq \lim_{n \rightarrow \infty} \liminf_{k \rightarrow \infty} \sum_{c \in \mathcal{X}^*} m(b, c) \hat{\nu}_{k,n}(a, c) \geq \lim_{n \rightarrow \infty} \sum_{c \in \mathcal{X}^*} m(b, c) \hat{\nu}_n(a, c) = \varpi(a, b),$$

for all $a, b \in \mathcal{X}$, where we have used the Fatou's Lemma in the last inequality.

Hence, there exists $(k(\varepsilon), n(\varepsilon)) \in \mathbb{N} \times \mathbb{N}$ such that $n \geq n(\varepsilon)$, $k > k(\varepsilon)$ implies

$$\hat{\varpi}_{k,n}(a, b) \geq \varpi(a, b) - \frac{\varepsilon}{2}, \text{ for all } a, b \in \mathcal{X}.$$

So, for large $k \geq k(\varepsilon)$ and large $n \geq n(\varepsilon)$ we have that

$$|\tilde{\varpi}_n(a, b) - \hat{\varpi}_{k,n}(a, b)| \leq |\tilde{\varpi}_n(a, b) - \varpi(a, b)| + |\varpi(a, b) - \hat{\varpi}_{k,n}(a, b)| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

by the triangle inequality. This gives

$$\lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \hat{\varpi}_{k,n}(a, b) = \lim_{n \rightarrow \infty} \tilde{\varpi}_n(a, b) = \varpi(a, b), \text{ for all } a, b \in \mathcal{X},$$

which concludes the proof of the Lemma. \square

Lemma 3.9 (Approximation Step 3). *Suppose (ϖ, ν) is sub-consistent. Then, for every $\varepsilon > 0$, there exists $k(\varepsilon)$ such that for large $k > k(\varepsilon)$, $|\varpi(a, b) - \varpi_k(a, b)| < \varepsilon$, for all $a, b \in \mathcal{X}$, $d(\nu_k, \nu) \leq \varepsilon$.*

Proof. We begin the proof, by noting that

$$\lim_{n \rightarrow \infty} \hat{\varpi}_{k,n}(a, b) = \varpi_k(a, b), \text{ for all } a, b \in \mathcal{X}$$

and $\lim_{n \rightarrow \infty} \hat{\nu}_{k,n} = \nu_k$ by construction. Fix $\delta = \frac{\varepsilon}{3}$, then there exists $(k(\varepsilon), n(\varepsilon)) \in \mathbb{N} \times \mathbb{N}$ such that $n \geq n(\varepsilon)$, $k \geq k(\varepsilon)$ implies

$$\begin{aligned} |\varpi_k(a, b) - \varpi_{k,n}(a, b)| &\leq \frac{3\delta}{2}, \\ |\varpi_{k,n}(a, b) - \varpi(a, b)| &\leq \frac{3\delta}{2}, \\ d(\nu_k, \hat{\nu}_{k,n}) &\leq \delta, \\ d(\hat{\nu}_{k,n}, \hat{\nu}_n) &\leq \delta, \end{aligned}$$

and

$$d(\hat{\nu}_n, \nu) \leq \delta.$$

Using the above inequalities we have

$$|\varpi_k(a, b) - \varpi(a, b)| \leq |\varpi_k(a, b) - \hat{\varpi}_{k,n}(a, b)| + |\hat{\varpi}_{k,n}(a, b) - \varpi(a, b)| \leq \frac{3\delta}{2} + \frac{3\delta}{2} = \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

Finally, we have that $d(\nu_k, \nu) \leq d(\nu_k, \hat{\nu}_{k,n}) + d(\hat{\nu}_{k,n}, \hat{\nu}_n) + d(\hat{\nu}_n, \nu) \leq 3\delta = 3(\frac{\varepsilon}{3}) = \varepsilon$. \square

Next we state a lemma analogous to Lemma 3.6 by Dembo et al [DMS03]. It will help us approximate pair of measures $(\varpi, \nu) \in \mathcal{M}_{c,k}$ with $\varpi_2 \neq \nu_1$ by a consistent $(\varpi_{x,y}, \nu_{x,y})$ with $(\varpi_{x,y})_2 = (\nu_{x,y})_1$, where $(\varpi_{x,y})_2$ and $(\nu_{x,y})_1$ are the second and the \mathcal{X} - marginals of the measures $\varpi_{x,y}$ and $\nu_{x,y}$, respectively. To do this, we review or collect some notation from [DMS03]. For $\nu \in \mathcal{M}_{c,k}$ and $a \in \mathcal{X}$ we write $\nu(\cdot, |a) = \nu(\cdot, a)/\nu_1(a)$ and

$$A_{0,0}(a, b) = \sum_{c \in \mathcal{X}_k^*} m(a, c) \nu(c|b), \text{ for } a, b \in \mathcal{X}.$$

As \mathcal{X}_k^* is finite we can find $b_0 \in \mathcal{X}_r$ such that $\mathbb{Q}_k\{c|a\} > 0$ and also $\nu(c|b) > 0$, such that $\sum_{a \in \mathcal{X}_r} m(a, c_2)$ is large enough to ensure that the difference $\sum_{a \in \mathcal{X}_r} \nu_{0,0}[m(a, c_2) - m(a, c_1(b))]$ > 0 . Let $c_1(b)$ be any number for $b \in \mathcal{X}_t$ and $c_2 = c_1(b)$ for all $b \neq b_0$. For any $|x| < 1/2$ we recall the definition of the probability measure $\nu_{x,0}$ as

$$\nu_{x,0}(c|b) = \nu(c|b) + x\nu(c_2|b)\nu(c_1|b)(1_{\{c=c_2\}} - 1_{\{c=c_1\}}).$$

Let $y_0 = \mathbb{Q}\{c_2|b_0\} \min_{b \in \mathcal{X}_r} \mathbb{Q}\{c_1|b\} > 0$, and for any $0 < y < y_0$ we recall the definition of the probability measures $\nu_{x,y}(\cdot|b)$ by

$$\nu_{x,0}(c|b) = \min(\nu_{x,0}(c_1|b), \mathbb{Q}\{c|b\}/y) \text{ for } c \neq c_1$$

$$\nu_{x,y}(c_1|b) = \nu_{x,0}(c_1|b) + \sum_{c \neq c_1} (\nu(c|b) - \mathbb{Q}\{c|b\}/y)_+$$

where $+$ indicates the positive part. Note that by construction

$$A_{x,y}(a, b) = \sum_{c \in \mathcal{X}^*} m(a, c) \nu_{x,y}(c|b) \rightarrow A_{0,0}(a, b), \text{ for any } a, b \in \mathcal{X}.$$

Finally, by $\varrho(A_{x,y})$ and $(\nu_{x,y})_1$ we denote Perron-Frobenius eigen value and eigen vector of the the weakly irreducible matrix $A_{x,y}$, respectively.

Lemma 3.10 (Approximation Step 4). *Suppose $(\varpi, \nu) \in \mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$ is consistent with $\varpi_2 \neq \nu_1$. Let $(a, c) \in \mathcal{X} \times \mathcal{X}_k^*$, then, for any $y \in (0, y_0)$ and $x(y) \in (-1/2, 1/2)$, we have*

- (i) $\nu_{x,y}(a, c) \leq \mathbb{Q}_k\{c|a\}(\nu_{x,y})_1(a)/y$,
- (ii) $\nu_{x,y}(a, c) \rightarrow \nu(a, c)$ as $x \rightarrow 0$ and $y \downarrow 0$,
- (iii) $\nu_{x,y}(a, c) = 0$ if and only if $\nu(a, c) = 0$,
- (iv) $\varpi_{x,y}(a, b) \rightarrow \varpi(a, b)$ as $x \rightarrow 0$ and $y \downarrow 0$, for all $b \in \mathcal{X}$.
- (v) $(\varpi_{x,y})_2(a) = (\nu_{x,y})_1(a)$

Moreover,

$$\limsup_{\substack{x \rightarrow 0 \\ y \downarrow 0}} \sum_{a \in \mathcal{X}} (\nu_{x,y})_1(a) H(\nu_{x,y}(\cdot|a) \| \mathbb{Q}_k\{\cdot|a\}) \leq \sum_{a \in \mathcal{X}} \nu_1(a) H(\nu(\cdot|a) \| \mathbb{Q}_k\{\cdot|a\})$$

The proof of this Lemma which is based on the Perron-Frobenius eigen value theorem and the implicit function theorem applied to the function $f(x, y) = \varrho(A_{x,y})$, is omitted. See, proof of [DMS03, Lemma 3.6].

For \mathbb{Q}_k we recall the definition of the rate function $J_k : \mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*) \rightarrow [0, \infty]$ from Theorem 2.3 as

$$J_k(\varpi, \nu) = \begin{cases} H(\nu \parallel \nu_1 \otimes \mathbb{Q}_k) & \text{if } (\varpi, \nu) \text{ is consistent and } \varpi_2 = \nu_1, \\ \infty & \text{otherwise.} \end{cases}$$

Lemma 3.11 below is a key ingredient in our proof of the lower bound in Theorem 2.1 and will be proved using the above four approximation Lemmas.

Lemma 3.11 (Approximation Step 5). *Suppose (ϖ, ν) is sub-consistent, $\varpi_2 = \nu_1$ and $\nu \ll \nu_1 \otimes \mathbb{Q}$. Then, for every $\varepsilon > 0$, there exists $(\hat{\varpi}, \hat{\nu}) \in \mathcal{M}_{c,k}$ such that $|\varpi(a, b) - \hat{\varpi}(a, b)| < \varepsilon$, for all $a, b \in \mathcal{X}$, $d(\nu, \hat{\nu}) \leq \varepsilon$, $\hat{\varpi}_2 = \hat{\nu}_1$ and*

$$J_k(\hat{\varpi}, \hat{\nu}) - J(\varpi, \nu) \leq \varepsilon.$$

Proof. Recall our assumption that (ϖ, ν) is sub-consistent, $\varpi_2 = \nu_1$ and $\nu \ll \nu_1 \otimes \mathbb{Q}$.

Case 1: If $(\varpi_k)_2 = (\nu_k)_1$ we take $\hat{\nu} := \nu_k$ and $\hat{\varpi} := \varpi_k$. Then, $\hat{\nu} \ll (\hat{\nu})_1 \otimes \mathbb{Q}_k$, and for every $\varepsilon > 0$, $|\varpi(a, b) - \hat{\varpi}(a, b)| < \varepsilon$, for all $a, b \in \mathcal{X}$, $d(\nu, \hat{\nu}) \leq \varepsilon$ and $\hat{\varpi}_2 = \hat{\nu}_1$ by Lemma 3.9. Now, using Lemma 3.6 we obtain

$$\limsup_{k \rightarrow \infty} J_k(\varpi_k, \nu_k) = \limsup_{k \rightarrow \infty} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \nu_k(a, c) \log \frac{\nu_k(a, c)}{(\nu_k)_1 \otimes \mathbb{Q}_k(a, c)} \leq \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \nu(a, c) \log \frac{\nu(a, c)}{\nu_1 \otimes \mathbb{Q}(a, c)} = J(\varpi, \nu).$$

Case 2: If $(\varpi_k)_2 \neq (\nu_k)_1$ we use Lemma 3.10 to choose $(\varpi_{k,x,y}, \nu_{k,x,y})$ with $(\varpi_{k,x,y})_2 = (\nu_{k,x,y})_1$. We take $\hat{\nu} := \nu_{k,x,y}$ and $\hat{\varpi} := \varpi_{k,x,y}$. Then, $\hat{\nu} \ll (\hat{\nu})_1 \otimes \mathbb{Q}_k$. Using Lemma 3.6 and Lemma 3.10 we obtain

$$\begin{aligned} \limsup_{k \rightarrow \infty} \limsup_{\substack{x \rightarrow 0 \\ y \downarrow 0}} J_k(\varpi_{k,x,y}, \nu_{k,x,y}) &= \lim_{k \rightarrow \infty} \limsup_{\substack{x \rightarrow 0 \\ y \downarrow 0}} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \nu_{k,x,y}(a, c) \log \frac{\nu_{k,x,y}(a, c)}{(\nu_{k,x,y})_1 \otimes \mathbb{Q}_k(a, c)} \\ &\leq \limsup_{k \rightarrow \infty} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \nu_k(a, c) \log \frac{\nu_k(a, c)}{(\nu_k)_1 \otimes \mathbb{Q}(a, c)} \\ &= J(\varpi, \nu), \end{aligned}$$

where $\nu_{k,x,y}(a, b) \rightarrow \nu_k(a, b)$ and $\varpi_{k,x,y}(a, b) \rightarrow \varpi_k(a, b)$ as $x \rightarrow 0$ and $y \downarrow 0$, for all $a, b \in \mathcal{X}$. This completes the proof of the lemma. \square

We recall that $C(v) = (N(v), X_1(v), \dots, X_{N(v)}(v))$ and note that, for every k such that $\min_{a \in \mathcal{X}} \mathbb{Q}\{\mathcal{X}_k^* | a\} > 0$ and any tree-indexed process x , we have that

$$\begin{aligned} \mathbb{P}\{X = x \mid |T| = n\} &\geq \mathbb{P}\{(X = x, C(v) \in \mathcal{X}_k^*, v \in V) \mid |T| = n\} \\ &= \prod_{v \in V, |T|=n} \mathbb{Q}\{\mathcal{X}_k^* | x(v)\} \times \mathbb{P}_k\{X = x \mid |T| = n\} \\ &\geq \left(\min_{a \in \mathcal{X}} \mathbb{Q}\{\mathcal{X}_k^* | a\} \right)^n \times \mathbb{P}_k\{X = x \mid |T| = n\}, \end{aligned} \tag{3.27}$$

where \mathbb{P}_k denote the law of tree-indexed process with initial distribution μ and offspring kernel \mathbb{Q}_k , and

$$\lim_{k \rightarrow \infty} \min_{a \in \mathcal{X}} \mathbb{Q}\{\mathcal{X}_k^* | a\} = \lim_{k \rightarrow \infty} \min_{a \in \mathcal{X}} \sum_{c \in \mathcal{X}_k} \mathbb{Q}\{c | a\} = \lim_{k \rightarrow \infty} \sum_{c \in \mathcal{X}^*} 1_{\{c \in \mathcal{X}_k\}} \min_{a \in \mathcal{X}} \mathbb{Q}\{c | a\} = 1,$$

since \mathcal{X} is a finite Alphabet. To complete the proof of the lower bound, we take $O \subset \mathcal{M}_s$. Then, for any $(\varpi, \nu) \in O$ sub-consistent with $\varpi_2 = \nu_1$, $\nu \ll \nu_1 \otimes \mathbb{Q}$ we may find $\varepsilon > 0$ with ball around (ϖ, ν) of radius 2ε contained in O . By our approximation Lemma 3.11, we may find $(\varpi_k, \nu_k) \in O \cap \mathcal{M}_{c,k}$ with $|\varpi_k(a, b) - \varpi(a, b)| \downarrow 0$, $d(\nu_k, \nu) \downarrow 0$, $(\varpi_k)_2 = (\nu_k)_1$, $\nu_k \ll (\nu_k)_1 \otimes \mathbb{Q}_k$ and

$$J_k(\varpi_k, \nu_k) - J(\varpi, \nu) \leq \varepsilon.$$

Hence, using the lower bound of Theorem 2.3 for offspring kernel \mathbb{Q}_k given by (3.20), (3.27) for large $k \geq k(\varepsilon)$ (with $\min_{a \in \mathcal{X}} \mathbb{Q}\{\mathcal{X}_k^* | a\} > 0$) and for large $n \geq n(\varepsilon)$, we obtain

$$\begin{aligned} \mathbb{P}\{(\tilde{L}_X, M_X) \in O \mid |T| = n\} &\geq \mathbb{P}\{|\varpi_k(a, b) - \tilde{L}_X(a, b)| < \varepsilon, \forall a, b \in \mathcal{X}, d(\nu_k, M_X) < \varepsilon \mid |T| = n\} \\ &\geq e^{n\alpha_k} \times \mathbb{P}_k\{|\varpi_k(a, b) - \tilde{L}_X(a, b)| < \varepsilon, \forall a, b \in \mathcal{X}, \\ &\quad d(\nu_k, M_X) < \varepsilon \mid |T| = n\} \\ &\geq \exp(-n(J_k(\varpi_k, \nu_k) + \varepsilon - \alpha_k)) \end{aligned}$$

where $\alpha_k = \log(\min_{a \in \mathcal{X}} \mathbb{Q}\{\mathcal{X}_k | a\})$. Taking limits we have that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{(\tilde{L}_X, M_X) \in O \mid |T| = n\} \geq -\lim_{k \rightarrow \infty} J_k(\varpi_k, \nu_k) - \lim_{k \rightarrow \infty} \alpha_k - \varepsilon \geq -J(\varpi, \nu) - 2\varepsilon.$$

Taking $\varepsilon \downarrow 0$ we have the desired result which completes the proof of the lower bound.

4. PROOF OF COROLLARIES 2.2, 2.5 AND THEOREM 2.4

4.1 Proof of Corollary 2.2. We derive this corollary from Theorem 2.1 by applying the contraction principle to the linear mapping $W : \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*) \mapsto \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ defined by

$$W(\varpi, \nu)(a, c) = \nu(a, c), \text{ for all } (a, c) \in \mathcal{X} \times \mathcal{X}^*.$$

In fact Theorem 2.1 implies the large deviation principle for $W(\tilde{L}_X, M_X)$ with convex, good rate function $\hat{J}(\nu) = \inf \{J(\varpi, \nu) : W(\varpi, \nu) = \nu\}$. Now, using sub-consistency and $\varpi_2 = \nu_1$ we obtain the form $\hat{J}(\nu) = H(\nu \| \nu_1 \otimes \mathbb{Q})$, for ν satisfying $\langle m(\cdot, c), \nu(a, c) \rangle \leq \nu_1$. Recall the definition of shift-invariant and denote by \mathcal{M}_1 set of shift-invariant measures in $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$. Write

$$\mathcal{M}_2 = \left\{ \nu : \nu \in \mathcal{M}(\mathcal{X} \times \mathcal{X}^*), \langle m(\cdot, c), \nu(a, c) \rangle \leq \nu_1 \right\}$$

and note that $\mathcal{M}_1 \subseteq \mathcal{M}_2$. Also, for all (values of) n where $\mathbb{P}\{|T| = n\} > 0$, we have

$$\mathbb{P}\{M_X \in \mathcal{M}_2 \mid |T| = n\} = 1.$$

Moreover, if $\nu_n \in \mathcal{M}_2$ converges to ν then

$$\nu_1(a) = \lim_{n \rightarrow \infty} (\nu_n)_1(a) \geq \liminf_{n \rightarrow \infty} \sum_{(b, c) \in \mathcal{X} \times \mathcal{X}^*} m(a, c) \nu_n(b, c) \geq \sum_{(b, c) \in \mathcal{X} \times \mathcal{X}^*} m(a, c) \nu(b, c),$$

which implies ν is sub-consistent. This means \mathcal{M}_2 is a closed subset of $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$. Therefore, by [DZ98, Lemma 4.1.5], the LDP for M_X conditional on the event $\{|T| = n\}$ holds with convex, good rate function K , which completes the proof of the corollary.

4.2 Proof of Theorem 2.4. We begin the proof of the theorem by stating the following Lemma.

Lemma 4.1. *Suppose that $q(c) = p(n) \prod_{i=1}^n \hat{q}(a_i)$ for all $c = (n, a_1, \dots, a_n)$, where $\hat{q}(\cdot)$ is a probability vector on \mathcal{X} and $p(\cdot)$ a probability measure with mean one on the nonnegative integers. Then, we have*

$$\inf \left\{ H(\tilde{\nu} \| q) : \tilde{\nu} \in \mathcal{M}(\mathcal{X}^*), \phi(b) = \sum_{c \in \mathcal{X}^*} m(b, c) \tilde{\nu}(c) \text{ for all } b \in \mathcal{X} \right\} = zH(\phi/z \| \hat{q}) + I_p(z), \quad (4.1)$$

where $\phi : \mathcal{X} \rightarrow \mathbb{R}$ and $z = \sum_b \phi(b)$.

Proof. For $\tilde{\nu} \in \mathcal{M}(\mathcal{X}^*)$, we let $\phi(b) = \sum_{c \in \mathcal{X}^*} m(b, c) \tilde{\nu}(c)$, for all $b \in \mathcal{X}$ and suppose first that $z = 0$, i.e. $\phi(b) = 0$ for all $b \in \mathcal{X}$. Then, $\tilde{\nu}((0, \emptyset)) = 1$ is the only possible measure in left side of (4.1), leading to $\tilde{I}(\phi, q) = -\log q((0, \emptyset)) = -\log p(0)$. It follows from (2.9) that $I_p(0) = -\log p(0)$ establishing (4.1) for such $\phi(\cdot)$. We assume hereafter that $z > 0$. Now the possible measures $\tilde{\nu}(\cdot)$ in the left side of (4.1) are of the form $\tilde{\nu}(c) = s(n)v_n(a_1, \dots, a_n)$ for $c = (n, a_1, \dots, a_n)$, with $v_0 = 1$, where $s(\cdot)$ is a probability measure on the nonnegative integers whose mean is z , and $v_n(\cdot)$, $n \geq 1$, are probability measures on \mathcal{X}^n with marginals $v_{n,i}(\cdot)$ such that

$$\phi(b) = \sum_{n=1}^{\infty} s(n) \sum_{i=1}^n v_{n,i}(b) \quad \text{for all } b \in \mathcal{X}. \quad (4.2)$$

By the assumed structure of $q(\cdot)$ we have for such $\tilde{\nu}(\cdot)$ that

$$H(\tilde{\nu} \| q) = \sum_{n=1}^{\infty} s(n) H(v_n \| \hat{q}^n) + H(s \| p), \quad (4.3)$$

where \hat{q}^n denotes the product measure on \mathcal{X}^n with equal marginals \hat{q} . Recall that

$$\sum_{n=1}^{\infty} s(n) H(v_n \| \hat{q}^n) \geq \sum_{n=1}^{\infty} s(n) \sum_{i=1}^n H(v_{n,i} \| \hat{q}) \geq zH\left(z^{-1} \sum_{n=1}^{\infty} s(n) \sum_{i=1}^n v_{n,i} \| \hat{q}\right),$$

with equality whenever $v_n = \prod_{i=1}^n v_{n,i}$ and $v_{n,i}$ are independent of n and i (see [DZ98, Lemma 7.3.25] for the first inequality, with the second inequality following by convexity of $H(\cdot \| \hat{q})$ and the fact that $\sum_n s(n)n = z$). So, in view of (4.2),

$$H(\tilde{\nu} \| q) \geq zH(\phi/z \| \hat{q}) + H(s \| p), \quad (4.4)$$

with equality when

$v_n = (z^{-1}\phi)^n$ for all $n \geq 1$. Now, write $\Lambda_p(\lambda) := \log \sum_n e^{\lambda n} p(n)$ and notice that Λ convex function and $\Lambda(0) = 0 < \infty$, and so, we have, for every $\lambda \in \mathbb{R}$, $\Lambda_p(\lambda) > -\infty$. Using Jensen's inequality, for every $s \in \mathcal{M}(\mathbb{N} \cup \{0\})$ and every $\lambda \in \mathbb{R}$, we have

$$\Lambda_p(\lambda) = \log \sum_n s(n) \left(\frac{e^{\lambda n} p(n)}{s(n)} \right) \geq \sum_n s(n) \log \left(\frac{e^{\lambda n} p(n)}{s(n)} \right) = \lambda \sum_n n s(n) - H(s \| p),$$

with equality if $s_\lambda(n) = p(n)e^{\lambda n - \Lambda(\lambda)}$. Thus, for all λ and all z , we have

$$\lambda z - \Lambda(\lambda) \leq \inf \{ H(s \| p) : s \in \mathcal{M}(\mathbb{N} \cup \{0\}) \text{ and } \sum_n s(n)n = z \} := \Lambda^*(z), \quad (4.5)$$

with equality when $\sum_n s(n)n = z$. Elementary calculus also shows that

$$\Lambda^*(z) = \lambda_* z - \Lambda_p(\lambda_*), \quad (4.6)$$

where λ_* is the solution of $\Lambda'_p(\lambda_*) = z$ and $\frac{d\Lambda_p}{d\lambda} := \Lambda'_p(\lambda)$. Combining (4.6) and (4.5) we obtain

$$\sup_{\lambda \in \mathbb{R}} \{\lambda z - \Lambda_p(\lambda)\} \leq \Lambda^*(z) \leq \sup_{\lambda \in \mathbb{R}, \Lambda'_p(\lambda)=z} \{\lambda z - \Lambda(\lambda)\} \leq \sup_{\lambda \in \mathbb{R}} \{\lambda z - \Lambda_p(\lambda)\}.$$

This yields $\Lambda^*(z) = I_p(z)$, which ends the proof of the Lemma. \square

Next, note that X is an irreducible, critical multitype Galton-Watson tree with offspring law

$$\mathbb{Q}\{c|b\} = p(n) \prod_{i=1}^n Q\{a_i|b\}, \text{ for } c = (n, a_1, \dots, a_n). \quad (4.7)$$

We derive Theorem 2.4 from Theorems 2.1 and 2.3 by applying the contraction principle to the continuous linear mapping $F : \mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*) \rightarrow \mathbb{R}^{\mathcal{X} \times \mathcal{X}}$, defined by

$$F(\varpi, \nu)(a, b) = \varpi(a, b), \text{ for all } (\varpi, \nu) \in \mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*) \text{ and } a, b \in \mathcal{X}. \quad (4.8)$$

It is easy to see that on $\{|T| = n\}$ we have $L_X = \frac{n}{n-1} F(\tilde{L}_X, M_X) = \frac{n}{n-1} \tilde{L}_X$. It follows that conditioned on $\{|T| = n\}$ the random variables L_X are exponentially equivalent to \tilde{L}_X , hence L_X satisfy the same large deviation principle as $F(\tilde{L}_X, M_X)$, see [DZ98, Theorem 4.2.13]. Without loss of generality we restrict the space for the large deviation principle of L_X to the set of all probability vectors on $\mathcal{X} \times \mathcal{X}$, see [DZ98, Lemma 4.1.5(b)].

Suppose p has finite second moment. Then, Theorem 2.1 implies the large deviation principle for $F(\tilde{L}_X, M_X)$ conditioned on $\{|T| = n\}$ with the good rate function $I(\mu) = \inf\{J(\mu, \nu) : F(\mu, \nu) = \mu\}$, see for example [DZ98, Theorem 4.2.1]. Convexity of I follows easily from the linearity of F and convexity of J .

Turning to the proof of (2.10), recall that ν is sub-consistent if and only if $F(\varpi, \nu)(a, b) \geq \sum_{c \in \mathcal{X}^*} m(b, c) \nu(a, c)$ for all $a, b \in \mathcal{X}$. Hence, we have that

$$I(\mu, \nu) = \inf \left\{ H(\nu \| \nu_1 \otimes \mathbb{Q}) : F(\mu, \nu)(\cdot, \cdot) \geq \sum_{c \in \mathcal{X}^*} m(\cdot, c) \nu(\cdot, c), \nu_1 = \mu_2 \right\}. \quad (4.9)$$

Note that $\nu_1(a) = 0$ yields $\sum_b F(\nu)(a, b) = 0$ if $F(\mu, \nu)(\cdot, \cdot) = \sum_{c \in \mathcal{X}^*} m(\cdot, c) \nu(\cdot, c)$ and

$$\sum_{(b, c) \in \mathcal{X}^*} m(b, c) \nu(a, c) < 0 \text{ if } F(\mu, \nu)(\cdot, \cdot) > \sum_{c \in \mathcal{X}^*} m(\cdot, c) \nu(\cdot, c).$$

Hence if $\mu_1(a) > 0 = \mu_2(a)$ for some $a \in \mathcal{X}$ then $\left\{ \nu : F(\mu, \nu)(\cdot, \cdot) = \sum_{c \in \mathcal{X}^*} m(\cdot, c) \nu(\cdot, c), \nu_1 = \mu_2 \right\} \cup \left\{ \nu : F(\varpi, \nu)(\cdot, \cdot) > \sum_{c \in \mathcal{X}^*} m(\cdot, c) \nu(\cdot, c), \nu_1 = \mu_2 \right\}$ is an empty set, and therefore $I(\mu) = \infty$. Assuming, throughout the rest of the proof that $\mu_1 \ll \mu_2$, it is not uneasy to verify that

$$I(\mu) = \sum_{a \in \mathcal{X}} \mu_2(a) \hat{I}\left(\frac{\mu(a, \cdot)}{\mu_2(a)}, \mathbb{Q}\{\cdot|a\}\right), \quad (4.10)$$

where for $q \in \mathcal{M}(\mathcal{X}^*)$, $\hat{I}\left(\frac{\mu(a, \cdot)}{\mu_2(a)}, q\right) = \inf \left\{ \tilde{I}(\phi, q) : \phi : \mathcal{X} \rightarrow \mathbb{R}_+, \phi(a) \leq \frac{\mu(a, \cdot)}{\mu_2(a)}, \text{ for all } a \in \mathcal{X} \right\}$ and

$$\tilde{I}(\phi, q) := \inf \left\{ H(\tilde{\nu} \| q) : \tilde{\nu} \in \mathcal{M}(\mathcal{X}^*), \phi(b) = \sum_{c \in \mathcal{X}^*} m(b, c) \tilde{\nu}(c) \text{ for all } b \in \mathcal{X} \right\}. \quad (4.11)$$

Suppose now that $q(c) = p(n) \prod_{i=1}^n \widehat{q}(a_i)$ for all $c = (n, a_1, \dots, a_n)$, where $\widehat{q}(\cdot)$ is a probability vector on \mathcal{X} and $p(\cdot)$ a probability measure with mean one on the nonnegative integers, whose second moment is finite. Then, by Lemma 4.1, we have the representation

$$\widetilde{I}(\phi, q) = \|\phi\| H(\phi/\|\phi\| \parallel \widehat{q}) + I_p(\|\phi\|), \quad (4.12)$$

where $\|\phi\| := \sum_{b \in \mathcal{X}} \phi(b)$. Therefore, it suffice for us to show that

$$\inf \left\{ \widetilde{I}(\phi, \widehat{q}) : \phi : \mathcal{X} \rightarrow \mathbb{R}_+, \phi(b) \leq \frac{\mu(a,b)}{\mu_2(a)}, \text{ for all } a \in \mathcal{X} \right\} = \frac{\mu_1(a)}{\mu_2(a)} H\left(\frac{\mu(a,\cdot)}{\mu_1(a)} \parallel \widehat{q}\right) + I_p\left(\frac{\mu_1(a)}{\mu_2(a)}\right). \quad (4.13)$$

To do this, we write

$$h(\phi(b)) := \widetilde{I}(\phi, \widehat{q}) + \alpha(b) \left(\phi(b) - \frac{\mu(a,b)}{\mu_2(a)} \right), \text{ for } b \in \mathcal{X},$$

where α is a Lagrange multiplier. Then, elementary calculus shows that $\alpha(b)$ is the solution of the equation

$$I_p\left(\frac{\mu_1(a)}{\mu_2(a)}\right) - \frac{\mu(a,b)}{\mu_2(a)} \sum_{a \in \mathcal{X}} e^{-\alpha(a)} \widehat{q}(a) = 0$$

and that $\phi(b) = \frac{\mu(a,b)}{\mu_2(a)}$ is the minimizer of our constraint optimization problem. Writing $\phi(b) = \frac{\mu(a,b)}{\mu_2(a)}$ in 4.12 we obtain left side of (4.13) which proves the theorem in case of p with unbounded support and finite second moments.

4.3 Proof of Corollary 2.5. Recall that T is Galton-Watson tree with offspring law $p(\ell) = 2^{-(\ell+1)}$, $\ell = 0, \dots$. Also, we recall that X is markov chain indexed by T with arbitrary initial distribution and transition kernel Q . Then, X satisfies all assumptions of Theorem 2.4, in particular we have $\sum_{\ell=0}^{\infty} \ell^2 p(\ell) = 3 < \infty$. Therefore, by Theorem 2.4, L_X conditioned on the events $\{|T| = n\}$ satisfies a large deviation principle in $\mathcal{M}(\mathcal{X} \times \mathcal{X})$ with good, convex rate function

$$I(\mu) = \begin{cases} H(\mu \parallel \mu_1 \otimes Q) + \sum_{a \in \mathcal{X}} \mu_2(a) I_p\left(\frac{\mu_1(a)}{\mu_2(a)}\right) & \text{if } \mu_1 \ll \mu_2, \\ \infty & \text{otherwise,} \end{cases} \quad (4.14)$$

where $I_p(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x + \log(2 - e^\lambda)\}$. Elementary Calculus shows that

$$\sup_{\lambda \in \mathbb{R}} \{\lambda x + \log(2 - e^\lambda)\} = x \log x - (x+1) \log \frac{(x+1)}{2}. \quad (4.15)$$

Therefore, writing (4.15) in (4.14) and rearranging terms we obtain the form of the rate function in the corollary which completes the proof.

Acknowledgments: I thank P. Mörters for helpful discussions on how to obtain large deviation lower bounds by conditioning, and the two referees (for my first submission) for many suggestions which have helped to improve the article.

BIBLIOGRAPHY

- [Al91a] D. ALDOUS. The continuum random tree I, In: *Ann. Probab.* 19, 1–28(1993).
- [Al91b] D. ALDOUS. The continuum random tree II: An overview. In: *Stochastic analysis*, Proc. Symp., Durham/UK 1990, Lond. Math. Soc. Lect. Note Ser. 167 (1991), 23–70.
- [Al93] D. ALDOUS. The continuum random tree III, In: *Ann. Probab.* 21, 248–289(1993).
- [AN72] K.B. ATHREYA AND P.E. NEY. Branching processes. Springer, New York, (1972).

- [BP94] I. BENJAMINI AND Y. PERES. Markov chains indexed by trees. *Ann. Probab.* 22 (1994), 219–243.
- [DA06] K. DOKU-AMPONSAH. Large deviations and basic information theory for hierarchical and networked data structures. PhD Thesis, Bath 2006.
- [DA10] K. DOKU-AMPONSAH. Asymptotic equipartition properties for hierarchical and networked structures. *ESAIM:Probability and Statistics*. DOI: 10.1051/ps/2010016 : Published online by Cambridge University Press: 03 February 2011.
- [DM10] K. DOKU-AMPONSAH AND P. MÖRTERS. Large deviation principle for empirical measures of coloured random graphs. *The annals of Applied Probability*, 20,6 (2010),1989-2021.
- [DMS03] A. DEMBO, P. MÖRTERS AND S. SHEFFIELD. Large deviations of Markov chains indexed by random trees. *Ann. Inst. Henri Poincaré: Probab.et Stat.*41, (2005) 971-996.
- [DZ98] A. DEMBO AND O. ZEITOUNI. Large deviations techniques and applications. Springer, New York, (1998).
- [DG02] T. DUQUESNE AND J.F. LE GALL, Random Trees, Levy Processes and spatial Branching Processes *Asterisque*, 281 (2002).
- [DU03] T. DUQUESNE A limit theorem for the contour process of conditioned Galton-Watson trees, *Ann. Probab.* 31, 996–1027(2003).
- [MM78] A. MEIR AND J.W. MOON. On the altitude of nodes in random trees. *Canad. J. Math.* 30 (1978), 997–1015.
- [OS02] P. OLOFSSON AND C.A. SHAW. Exact sampling formulas for multitype Galton-Watson processes. *Journal of Mathematical Biology* (2002) **45**, 279-293.
- [KI11a] I. KORTCHEMSKI Random stable laminatrion of the disk. *arvix* 1106.0271(2011).
- [KI11b] I. KORTCHEMSKI Invariance principles for conditioned Galton-Watson trees. *arvix* 1110.2163v1.[math.PR] 10 Oct 2011.
- [Pe95] R. PEMANTLE. Tree-indexed processes. *Statist. Sci.* 10 (1995), no. 2, 200–231.
- [RD11] D. RIZZOLO. S.caling limits of Markov branching trees and Galton-Watson trees conditioned on nthe number of vertices with out-degree in a given set. *arxiv*: 1105 2528(2011).